TAP
Technology Advisory Panel
IPCO

Proportionality aide memoire: Privacy intrusion in data collection and analytics

24 November 2022

This aide memoire deals specifically with factors to consider in proportionality and assumes the cases for necessity and resources have already been made. Factors in bold and starred(*) usually carry more weight.

1. Factors relevant to data collection and analytics

Value	
Timeliness and need *	• gravity and extent of (potential) crime or harm • public interest • urgency of need
Function *	• for analysis of the data on its own • to enrich existing data • to become enriched by existing data • for training sets for use in machine learning algorithms in established tools • for use in development or enhancement of a new capability or tool, which may be a prototype
Relevance and marginal benefits *	• to given investigation(s) • to other data available
Impact of time and place	• dependencies such as when and where data were collected
Type of data or collection method	• new or existing type of data • new, more accurate, or existing collection method

Volume	
Amount *	• fixed and known before collection • unknown but can be approximated • granularity and uncertainties of approximations including dependencies
Frequency	• one-time collection • repeated collection, how many times and at which intervals • continuous collection, for how long • how does the amount of data held vary over time

Data Management	
Storage	• where, how, and under whose authority • length of time planned retention, for which parts • security of access and resilience to loss or corruption
Deletion and manipulation	• plans and mechanisms for indexing, deletion and/or putting beyond use, redaction, and abstraction

Analysis	
Human and/or machine inspection (*)	• uncertainty (false positives/negatives) thresholds for human and machine inspection • risks of bias for human and machine inspection • human only inspection is possible of entire data set • machine only inspection is possible of entire data set • primary analysis by machine inspection to extract set for secondary analysis by human inspection

Alternatives	
What other methods have been considered	• if they have been implemented successfully, why are they not employed now • if they have not been implemented successfully, why not • opportunity cost - what will be lost by implementing this method over others • efficiency and effectiveness of proposed method vs. alternatives

2. Factors relevant to intrusiveness

Privacy Intrusion	
Type (*)	• degrees of foreseeable, targeted, collateral, and privileged intrusion – how many individuals • their interrelationships and dependencies
Sensitivity (*)	• degree of sensitivity of the data collected and/or what will be revealed through subsequent analytics
Scaling	• how the intrusion scales from individuals to different populations e.g. multiplicative, additive, constant • how the intrusion affects a community defined by a characteristic
Access	• breadth of people (e.g. analysts) and systems that will have access either directly to the data collected or indirectly via analytical tools • breadth of people (e.g. analysts, colleagues, managers) who will have access to reports that refer to the data